**15th International Roundtable on Business Survey Frames**

**Washington, D.C. – October 22 – 26, 2001**

**Session No 5**
**Paper No 2**
**Colleen Porteous, Statistics New Zealand, New Zealand**

# The Quest for Quality – a Geocoding Quality Investigation

## 1.0 Introduction

*The aim of this paper is to:*

- promote the sharing of information about the current and the potential uses of geocoding business frames.

*To meet this aim, it was necessary to:*

- identify and explain the current geocoding practices used by Statistics New Zealand's (SNZ) Business Frame.
- assess the existing quality of geocoding on the SNZ Business Frame

Since 1998, the Business Frame Section has carried out a continuous programme of quality investigations. The results of the earlier stages of this programme were reported to the two previous Roundtables in the papers entitled "*The Quest for Quality - Establishing Industrial Classification, Life Cycle and Coverage Measurements in Statistics New Zealand's Business Frame"* and "*Challenges in the Quest for Quality – Lifecycle Coding Case Studies".* Today's presentation will outline the results of the final part of this programme: The investigation into the quality of geocoding on SNZ's Business Frame.

## 1.1 What is Geocoding?

Geocoding is the process of creating geographic co-ordinates for geographically referenced tabular data. In other words, a geocoding process will allow the derivation of precise co-ordinates on the surface of the earth for things like physical addresses. Because addresses are the geographic identifier for many databases or registers, an address can be matched to data in a Geographic Information System (GIS)[1] and a geographic location or co-ordinate can be assigned.

Generally there are two ways in which geocoding can be employed for the collection and presentation of statistical information. The first of these methods is a polygon-based approach where statistics for small and large scale areas are linked to polygons overlaid onto a topographic background map. The second method involves the use of a point-based approach where the statistical information is linked to x/y co-ordinates.

Statistics New Zealand (SNZ) uses a polygon based approach, the result of which is a basic spatial unit (BSU) termed a 'meshblock'. Geocoding using a meshblock allows SNZ to publish social and business information from censuses and surveys, by specified geographic area.

## 1.2 Background

New Zealand is divided geographically into 16 regional areas. These are then subdivided into the following :

[1] Geographical Information Systems are an integrated set of software tools for the collection storage, integration and analysis of geographically located data.

**Territorial Authorities (TA).** There are 74 TAs in New Zealand. TAs are a collection of Area Units which generally define significant areas of the country and population, specifically 15 cities and 59 districts.

**Area Units (AU).** Area Units are aggregations of meshblocks. They are non-administrative areas intermediate between meshblocks and territorial authorities. Area Units are single geographic entities with a unique meaningful name referring to a geographic feature. They contain an average of 5000 persons but can vary in terms of size and population. Urban area units are roughly equivalent to a suburb or part thereof.

**Meshblock (MB).** The meshblock is the smallest spatial unit used by SNZ in the collection and/or processing of data. It is a required field on SNZ's Business Frame and is the building block for aggregation into larger areas such as area units, and territorial authorities. (refer Appendix 1)

***Table 1***

| | North Island | South Island (incl Chathams Isl) | New Zealand Total |
|---|---|---|---|
| **Meshblocks** | 27337 | 10735 | 38072 |
| **Area Units** | 1243 | 560 | 1803 |
| **Territorial Authorities** | 49 | 25 | 74 |
| **Regional Councils** | 9 | 7 | 16 |

### 1.3 Uses of the Meshblock

The meshblock is used for :

- enumerating, classifying and disseminating geographic information;
- linking statistical (geographic) units to higher order geographic classifications;
- defining administrative, electoral and statistical boundaries

As well as producing regional data from individual surveys, the meshblock-coded geographic unit enables SNZ to link place of work statistics from the population census with regional data from administrative surveys that are linked to the business frame.

Regional information is also used in conjunction with socio-economic data for:

- locating resources (e.g. labour, raw materials).
- locating markets (e.g. retail planning).
- transport and routing applications (e.g. product delivery, journey to work studies)
- civil defence and emergency planning.
- internal migration studies.

*Refer Appendices for more detail*

### 1.4 Limitations of the Meshblock

- **Confidentiality Issues**: The meshblock was originally developed for use in population statistics and was never intended to be used for dissemination purposes. Since many meshblocks contain no businesses, or only one or two, meshblocks must often be aggregated to higher units to preserve confidentiality. For this reason, most information is provided at an aggregated level, or concorded to AUs or TAs.

- **Maintenance Issues**: The meshblock pattern is regularly being updated to reflect changes in administrative boundaries, new road patterns and new subdivisions. Data within affected

meshblocks must be rebased for time series analyses. This can be time consuming and may degrade data quality. (refer Appendix 1)

- **Conceptual Issues**: Use of the meshblock as a basic spatial unit has its own set of problems. It is conceptually less useful for business statistics because the meshblock does not take into account the spatial pattern of economic activity:
  - External clients often require SNZ data aligned with their own boundary pattern (e.g. sales area catchments), which may cut across meshblock boundaries. Since statistics cannot be disaggregated beyond the meshblock level, the data supplied by SNZ may not be exactly what clients want.
  - In some cases, businesses (especially farms or forests) may straddle more than one meshblock resulting in classification problems.
  - Many businesses carry out their activities away from their base location, e.g. fishermen, builders, taxi drivers. Often there is no consistent physical location for most of their activities. SNZ codes these businesses to the base location, except in the case of fishing, where the activity is coded to the nearest 'wet' area to the business base.

- **Operational Issues**: In order to code a meshblock accurately, a specific physical address is required. The major limitation therefore is that accuracy levels are based on the quality of the addresses provided.
  - Unnumbered addresses (with only a road name, but no number) have potential matches with more than one meshblock area. It is often not possible to provide an accurate meshblock code automatically. Thus, any measurement of the quality of meshblock coding must be assessed at both the rural and urban levels, with urban areas generally being able to be meshblocked to a higher degree of accuracy.
  - Corner addresses are also a problem, when 2 or more (often 4) meshblocks may meet at a corner. Respondents also supply incorrect numbers (or no numbers at all) which may not match to a single meshblock area.
  - Another problem is insufficient locality information on an address. For instance there are respondents who provide only suburb names, and no city. Where the same road name occurs in the same named suburb in different TAs (eg Main Road, Richmond in three different TAs) again coding is extremely difficult.

For the above reasons it is unlikely that a polygon based system would ever achieve 100% accuracy and therefore it is highly unlikely that we would ever get 100% match-rate for an address list.

**2.0 Quality of Geocoding on SNZ's Business Frame**

As a result of the work done for the Frame Quality Measurement Project 1998-1999 the Business Frame section set up a programme to investigate and report on the quality of BF coverage and of selected variables. The following studies are now available:

- **Coverage** of the Business Frame
- **Accuracy** of the Industry Classification Code
- The **Timeliness** of the Lifecycle Code
- **Quality** of Business Type and Institutional Sector coding on the BF
- **Quality** of **Geocoding** on the Business Frame

These investigations were designed to measure the quality of the most important attributes on the Business Frame. This information is important when determining what resources should be applied to frame maintenance and where these resources should be directed, particularly in the context of competition within the statistical agency for scarce resources. The final investigation regarding meshblock quality is reported here.

### 2.1 Objectives

The objectives of the BF Geocoding Quality Investigation were to:

- assess the quality of meshblock coding on SNZ's Business Frame;
    - identify what, if any differences existed between urban & rural quality of coding ;
    - investigate the quality of the hierarchical structure from Meshblock to Area Unit to Territorial Authority using direct concordances.
    - assess the impact of the quality of meshblock coding on selected strata levels
- identify factors affecting the performance of the geocoding process.
- propose possible remedial strategies

### 2.2 Methodology

A stratified random sample for investigating the meshblock quality was extracted based on the advice of our Survey Methodology division. The sample below (Table 2) was selected from the Business Frame population of economically significant enterprises :

***Table 2***

| | Geographic Units | |
|---|---|---|
| | **URBAN** | **RURAL** |
| **Greater than $1,000,000 or part of a group or multi geo** | *Urban_1* **500** | *Rural* **500** |
| **Greater than $100,000 less than $1,000,000 - Single geo not part of a group** | *Urban_2* **500** | |
| **Less than $100,000 - Single geo not part of a group** | *Urban_3* **500** | |

The sample selected represents the population numbers in Table 3:

***Table 3***

| | Geographic Units | |
|---|---|---|
| | **URBAN** | **RURAL** |
| **Greater than $1,000,000 or part of a group or multi geo** | *Urban_1* **70,316** | *Rural* **82,146** |
| **Greater than $100,000 less than $1,000,000 - Single geo not part of a group** | *Urban_2* **113,419** | |
| **Less than $100,000 - Single geo not part of a group** | *Urban_3* **106,245** | |

The first stage of the project involved the coding of the sample using SNZ's standard meshblock geocoder. For this study, the scoring level of the coder was raised to 100% single match. This is a higher standard than normally required in order to eliminate any possibility of geocoder/updater error within the study. (refer Appendix 2). Those remaining addresses were then coded manually to the same standard using all available tools. The two most common geographic coding tools available to staff are the geocoder and Terraview[2]. However, other resources such as the NZ Companies Office registrations, telephone guides and electoral roll information were used to clarify address information so that a meshblock could be assigned.

### 2.3 Factors affecting the Performance of the Geocoding Process

In any coding situation, the quality of the information supplied will always affect the accuracy of the resulting classifications. The primary reason for units failing to be automatically geocoded was because the information supplied was imprecise at some level. To better understand the impact of any imprecision on the quality of geocoding, the units which required manual updating were analysed further. As these units were coded, the manual updater made a note as to why the address would have caused the automatic process to fail. In general, the most difficult addresses to geocode were those without a site number and where occupiers were not the owners of the property. However there were other idiosyncratic problems as outlined below:

***Figure 1***

- Multiple matches due to *inadequate address* information.

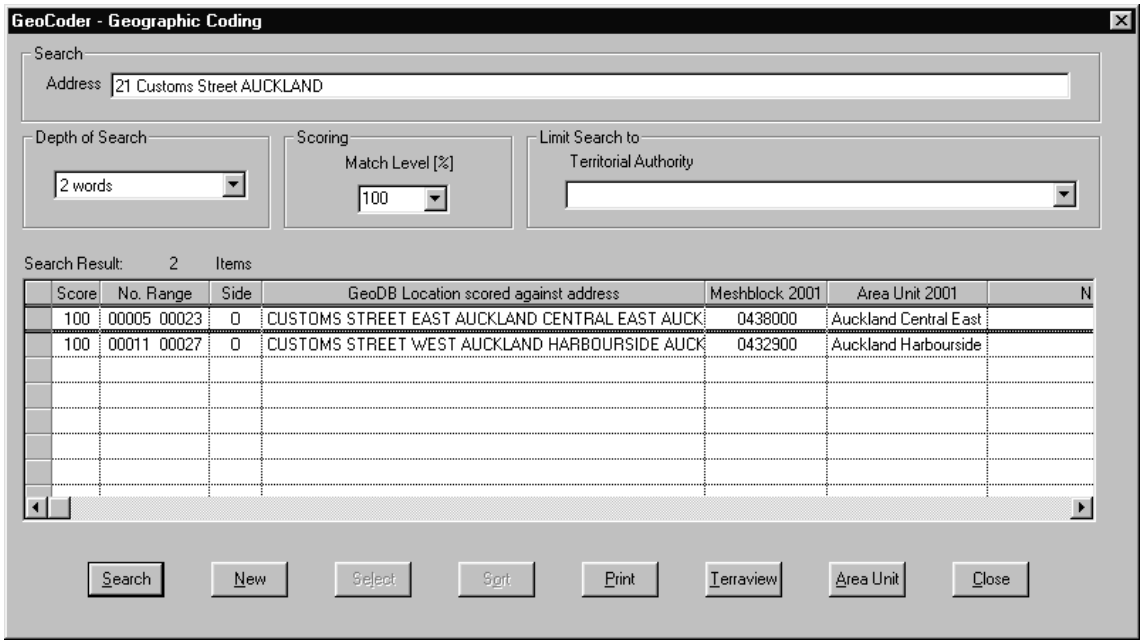



Figure 1 above is an example of an address which has been supplied with a correct number. However, the number & street name match to more than one street and thus the coder matches to both addresses correctly at 100%. In this case, a final coding decision is required by a manual updater.

[2] Terraview ™ is a Geographic Information System (GIS) data viewer software package which is designed to operate within users' offices either as a stand-alone on a Windows configured workstation or in a networked situation. It provides fast and easy access to land related information and can be used as a mapping base for other activities. It also provides valuable leads for valuation and local authority information. Terraview allows searching by a number of fields, including street address, land owner, lot and plan number, and geographic location.

- Addresses on the BF did not always reflect recent improvements in rural numbering via the Rural Address Property Identification System (RAPID)[3].

- BF units often presented difficulties with corner site addresses (e.g. “Oxford Tce & Worcester St”), and with site numbers which fell within an address range (e.g. “11-17 Adelaide St”).

***Figure 2***

- Multiple matches due to *inadequate numbering*:

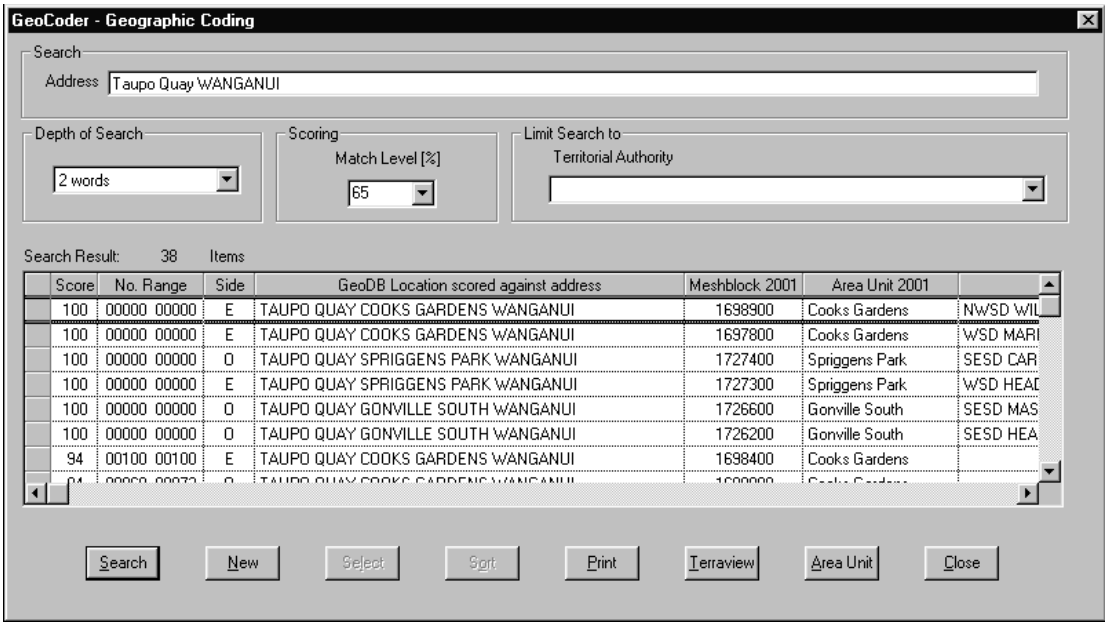



Figure 2 above shows the scoring level for the address ‘Taupo Quay, WANGANUI’. The geocoder has scored 6 possibilities at 100%, all with different meshblocks. The lack of a numbered address makes pinpointing difficult and thus returns multiple results.

Those meshblocks which required manual processing (coded manually) were sorted into four categories as follows:

- **Inadequate Numbering**. The geocoder was unable to code because of incorrect or missing street numbers.

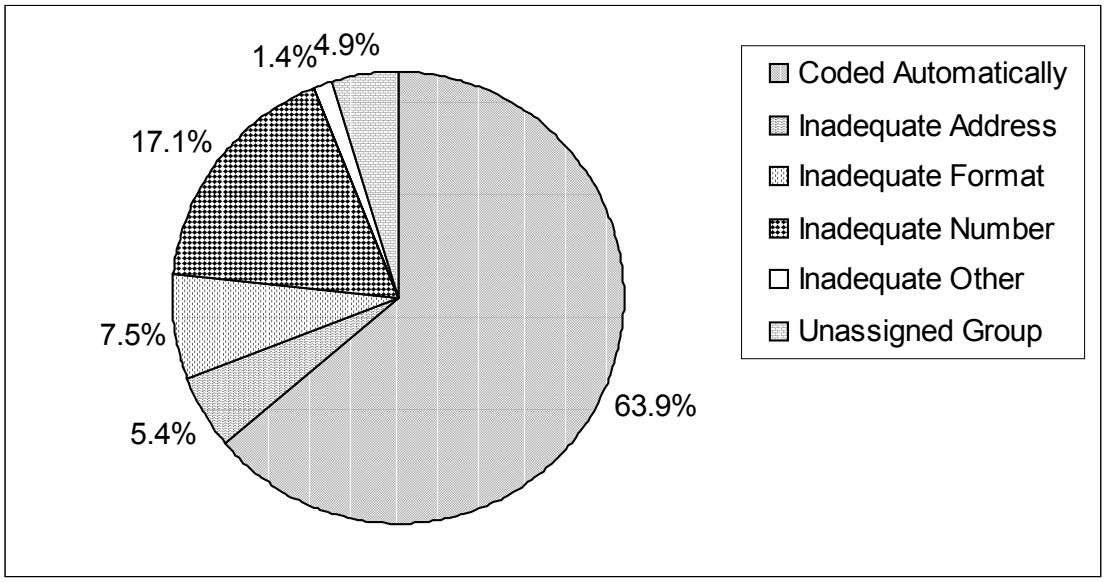
- **Inadequate Format**. This refers to the difference between real world versions of addresses and the more technically correct version required by the geocoder. (eg Sandringham as opposed to Sandringham East)

- **Inadequate Address**. The location address supplied not a recognised location on the geocoder – usually occurs when respondent says they live on ‘Main Rd’ when in fact this road has another name

- **Inadequate Other**. Units which did not fit the other categories easily or where a combination of factors caused the geocoding to fail to meet the standards of the investigation

[3] The Rural Address Property Identification System(RAPID) is a numbering system being implemented by local authorities, designed to aid emergency response times for rural areas – effectively a combination of a street numbering system with some relation to the distance travelled from a specific landmark

***Chart 1***




These four categories plus:

- **Coded Automatically**. where the geocoder was able to assign one meshblock at the 100% score level;

**and**

- **Unassigned Group**. where neither the geocoder nor a manual updater were able to assign a meshblock with a 100% confidence of accuracy.

were further analysed by strata levels.

***Table 4***

| | *Coded Automatically* | *Coded Manually* | | | | *Unassigned Group* | *Grand Total* |
|---|---|---|---|---|---|---|---|
| *STRATA* | | *Inadequate Address* | *Inadequate Format* | *Inadequate Number* | *Inadequate Other* | | |
| *Rural* | 154 (31%) | 50 (10%) | 34 (7%) | 190 (38%) | 7 (1%) | 65 (13%) | **500** |
| *Urban_1* | 350 (70%) | 22 (4%) | 46 (9%) | 65 (13%) | 6 (1%) | 11 (2%) | **500** |
| *Urban_2* | 382 (76%) | 15 (3%) | 38 (8%) | 45 (9%) | 7 (1%) | 13 (3%) | **500** |
| *Urban_3* | 392 (78%) | 20 (4%) | 31 (6%) | 42 (8%) | 7 (1%) | 8 (2%) | **500** |
| *Grand Total* | **1278 (64%)** | **107 (5%)** | **149 (8%)** | **342 (17%)** | **27 (1%)** | **97 (5%)** | **2000** |

The interesting point to note from this table is the high incidence of inadequate numbering (IN) for rural addresses as a source of coding difficulty. Whilst 17% of the overall difficulty was attributed to inadequate numbering, this increased to 38% for the rural component.

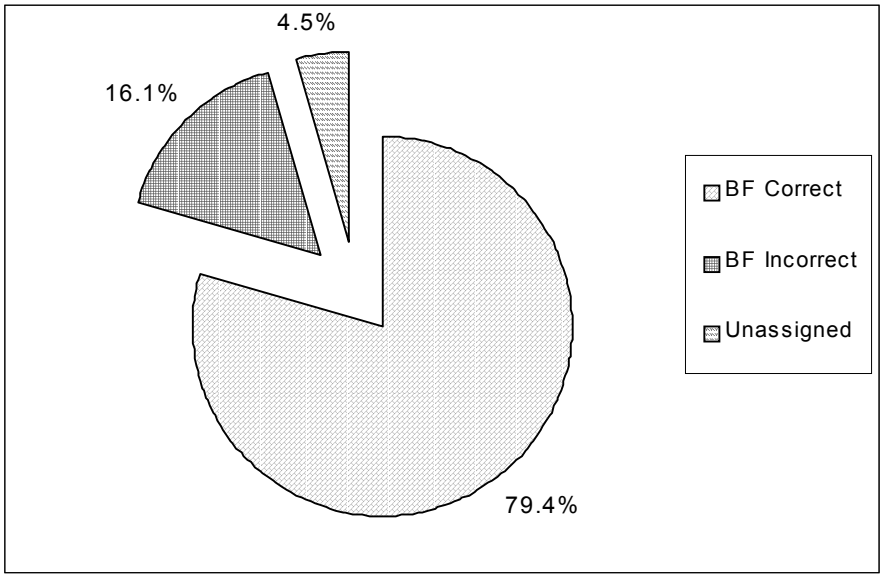
The other area in which rural coding showed significant areas of difficulty is the Inadequate Address category showing 10% of rural addresses lacked sufficient information to code automatically whereas the three urban strata are between 3-4%.

### 2.4 Results of the Investigation

A meshblock code, whether provided automatically or manually, was assigned to the address only if the meshblock given could be guaranteed to be 100% correct. This meshblock was then compared with the existing meshblock on the Business Frame to determine the accuracy of current geocoding practices

- **79.4%** were correct on the Business Frame. (Table 5)
- **16.1%** were incorrect on the Business Frame. (Table 6)
- **4.5%** could not be confidently assigned a meshblock for the investigation.(Table 7)

***Chart 2***




### 2.5 Analysis by Spatial Unit

### *Meshblock Results*

By analysing the resulting dataset by strata, some conclusions could be drawn as to the overall quality of the meshblock and the varying levels of quality between urban and rural meshblocking.

***Table 5 – Meshblock Correct***

| | Pop Size | Sample Size | # correct | weighted # correct | weighted percent correct |
|---|---|---|---|---|---|
| **Rural** | 82146 | 500 | 284 | 46659 | 56.8 |
| **Urban_1** | 70316 | 500 | 430 | 60472 | 86.0 |
| **Urban_2** | 113419 | 500 | 429 | 97314 | 85.8 |
| **Urban_3** | 106245 | 500 | 428 | 90946 | 85.6 |
| **Total** | 372126 | 2000 | 1571 | 295390 | 79.4 |

| 79.4 | **percent correct** |
|---|---|

Table 5[4] shows the number of meshblocks which matched to the BF meshblock code. These results were then weighted using stratum weights to reflect their proportions on the BF. This table clearly

[4] Eg Percent Correct = $\frac{\text{Weighted \# Correct}}{\text{Pop Size}} \times 100$

shows that while those units in the rural strata were correct only 57% of the time, in all three urban strata, the accuracy level is above 85%. The overall **correctly BF coded** rate is 79.4%.

***Table 6 – Meshblock Incorrect***

| | Pop Size | Sample Size | # incorrect | weighted # incorrect | weighted percent incorrect |
|---|---|---|---|---|---|
| **Rural** | 82146 | 500 | 151 | 24808 | 30.2 |
| **Urban_1** | 70316 | 500 | 59 | 8297 | 11.8 |
| **Urban_2** | 113419 | 500 | 58 | 13157 | 11.6 |
| **Urban_3** | 106245 | 500 | 64 | 13599 | 12.8 |
| **Total** | 372126 | 2000 | 332 | 59861 | 16.1 |

16.1 **percent incorrect**

Table 6 shows the proportions of **incorrectly BF coded** meshblocks over the four strata. Again, evident from this is the lack of quality of the rural meshblock. 30.2% of rural meshblocks were incorrect yet on average only 12% were incorrect in the urban strata.

***Table 7 – Meshblock Unassigned***

| | Pop Size | Sample Size | # unassigned | weighted # unassigned | weighted percent unassigned |
|---|---|---|---|---|---|
| **Rural** | 82146 | 500 | 65 | 10679 | 13.0 |
| **Urban_1** | 70316 | 500 | 11 | 1547 | 2.2 |
| **Urban_2** | 113419 | 500 | 13 | 2949 | 2.6 |
| **Urban_3** | 106245 | 500 | 8 | 1700 | 1.6 |
| **Total** | 372126 | 2000 | 97 | 16875 | 4.5 |

4.5 **percent unassigned**

Table 7 shows the number of units which were **unable to be confirmed as a correctly coded** meshblock. This does not imply that these units do not have meshblocks on the business frame but rather that, in the course of this investigation, neither the geocoder nor a manual updater were able to assign a meshblock to the **100% confidence requirement of this study.**

**2.6 Further Analysis by Published Spatial Units**

The BF geocoder is able to automatically concord a meshblock code to its Area Unit (AU) and Territorial Authority (TA) codes. Using the concordances available, the assigned meshblocks were then concorded to their AU and TA values. This enabled a further measurement of the quality of these two broader, but more widely published spatial units.

**Note***: When a meshblock did not meet the 100% confidence criteria, it could not be included in this concordance process. Thus the unassigned section (4.5%) remains excluded from these categories. Therefore the AU and TA percentages which have been 'upwardly derived' understate the likely accuracy rate which would have existed if units had been coded directly at these levels.*

***Area Unit Results***

***Table 8 – Area Unit Correct***

| | Pop Size | Sample Size | # correct | weighted # correct | weighted percent correct |
|---|---|---|---|---|---|
| Rural | 82146 | 500 | 387 | 63581 | 77.4 |
| Urban_1 | 70316 | 500 | 469 | 65956 | 93.8 |
| Urban_2 | 113419 | 500 | 458 | 103892 | 91.6 |
| Urban_3 | 106245 | 500 | 453 | 96258 | 90.6 |
| Total | 372126 | 2000 | 1767 | 329687 | 88.6 |

88.6 **percent correct**

Table 8 shows that the overall quality value of 88.6% at the Area Unit level is substantially higher than for the Meshblock (79%). The most notable gain made in this area is in the rural strata where accuracy levels have increased by approximately 21 percentage points, compared to urban units, which increased their level of accuracy by 5-8 percentage points.

***Table 9 – Area Unit Incorrect***

| | Pop Size | Sample Size | # incorrect | weighted # incorrect | weighted percent incorrect |
|---|---|---|---|---|---|
| Rural | 82146 | 500 | 48 | 7886 | 9.6 |
| Urban_1 | 70316 | 500 | 20 | 2813 | 4.0 |
| Urban_2 | 113419 | 500 | 29 | 6578 | 5.8 |
| Urban_3 | 106245 | 500 | 39 | 8287 | 7.8 |
| Total | 372126 | 2000 | 136 | 25564 | 6.9 |

6.9 **percent incorrect**

Table 9 shows that the number and proportions of incorrect Area Units has dropped also. Again those area units derived from rural meshblocks have seen the biggest improvement in quality by around 22 percentage points.

***Territorial Authority Results***

***Table 10 - Territorial Authority Correct***

| | Pop Size | Sample Size | # correct | weighted # correct | weighted percent correct |
|---|---|---|---|---|---|
| Rural | 82146 | 500 | 424 | 69660 | 84.8 |
| Urban_1 | 70316 | 500 | 488 | 68628 | 97.6 |
| Urban_2 | 113419 | 500 | 485 | 110016 | 97.0 |
| Urban_3 | 106245 | 500 | 491 | 104333 | 98.2 |
| Total | 372126 | 2000 | 1888 | 352637 | 94.8 |

94.8 **percent correct**

Table 10 clearly illustrates the improvements able to be achieved if coding was required at this broader level only, showing that over all 4 strata, the TA held on the business frame was correct 95% overall. This figure makes more sense when you add the 4.5% of missing values which leaves 0.7% of units 'incorrect'. (Table 11)

***Table 11 - Territorial Authority Incorrect***

| | Pop Size | Sample Size | # incorrect | weighted # incorrect | weighted percent incorrect |
|---|---|---|---|---|---|
| **Rural** | 82146 | 500 | 11 | 1807 | 2.2 |
| **Urban 1** | 70316 | 500 | 1 | 141 | 0.2 |
| **Urban 2** | 113419 | 500 | 2 | 454 | 0.4 |
| **Urban 3** | 106245 | 500 | 1 | 212 | 0.2 |
| **Total** | 372126 | 2000 | 15 | 2614 | 0.7 |

0.7 **percent incorrect**

**2.7 Summary of Results:**

***Meshblock***

- **79.4%** were correct on the Business Frame.
- **16.1%** were incorrect on the Business Frame.
- **4.5%** could not be confidently assigned a meshblock for the investigation.

***Area Unit ****

- **88.6%** were correct on the Business Frame.
- **6.9%** were incorrect on the Business Frame.

***Territorial Authority ****

- **94.8%** were correct on the Business Frame.
- **0.7%** were incorrect on the Business Frame.

***Table 12***

| | Meshblock | Area Unit | Territorial Authority |
|---|---|---|---|
| ***Rural*** | 56.8% | 77.4% | 84.8% |
| ***Urban*** | 85.8% | 91.8% | 97.6% |
| ***Total*** | 79.4% | 88.6% | 94.8% |

* Due to our programs producing area unit (AU) and territorial authority (TA) information directly from a given meshblock code, we were unable to assign either an AU or TA if the meshblock was missing.

**2.8 Estimated effect on survey outputs:**

The results of any BF quality study should ultimately be measured in terms of the effects on survey outputs. In the case of the meshblock study an analysis was undertaken to determine the effects on

the quality of SNZ's employment size measure, Full Time Equivalent Persons Engaged (FTE)[5]. The FTE numbers for the *population*, *the sample* and *those correctly coded in the sample* (for each strata) were extracted. Weighted percentages were then produced to reflect the accurate proportions of correctly coded employment statistics across the three geographic levels.

***Table 13***

| | Meshblock | Area Unit | Territorial Authority |
|---|---|---|---|
| ***Rural*** | 54.9% | 72.6% | 83.1% |
| ***Urban*** | 84.8% | 92.1% | 98.3% |
| ***Total*** | 81.6% | 90.0% | 96.7% |

The figures in Table 13 show that the results obtained from the investigation are consistent with the published FTE figures. The totals show that the overall quality of the FTE statistic is approximately 2 percentage points higher than the overall meshblock quality level. However, it is interesting to note that the rural population does not follow this trend. While the team has not undertaken an in-depth analysis into the reasons for this, the probable explanation is that this discrepancy is a result of two factors:

- there is a significantly smaller proportion of employing businesses in the rural strata (refer Appendix 3)
- non employing and agricultural businesses are not currently maintained by SNZ's Annual Frame Updating Survey

*(Note: Employment is not a reliable measure of the size of rural businesses.)*

**2.9 Recommendations:**

- That the Frame Updating Questionnaires be redesigned to take advantage of changes in postal technologies and facilitate the capture of new address information eg. the RAPID initiative;
- That the geocoder system should be reviewed to ensure that all available sources of information are being effectively utilised;
- That staff training should be widened to include a broader understanding of the role and importance of the meshblock to SNZ outputs;
- That both internal and external clients be fully apprised of the results of the Frame Quality Measurement Programme

**3.0 The future of geocoding in SNZ**

The growing need for more precise geographic information has lead to a debate within SNZ as to whether the meshblock best meets current and future user requirements. The question for the future of geocoding data within SNZ is: do we retain the meshblock as our lowest level spatial unit?

**Competitive Advantage**
Emerging trends overseas reveal two important, and connected developments; the increasing popularity of desktop Geographic Information Systems (GIS), and a growth in applications using socio-economic and business statistics. High growth areas include sales and marketing, real estate, banking, and transport management. The rapidly increasing power-to-price ratio of GIS technology has

[5] Full-time equivalent persons engaged (FTE) equals the sum of full-time employees and working proprietors plus half the part-time employees and working proprietors.

reduced the entry costs for smaller companies wanting to analyse spatial information and has increased demand for relevant datasets. This has important implications for SNZ. First, demand for business statistics integrated with spatial data is likely to increase. As clients become more exposed to digital datasets, expectations regarding SNZ's ability to deliver digital data will increase. Also, since data acquisition is a major cost for any new GIS system, the provision of reasonably priced integrated digital datasets would open up new markets. SNZ needs to be in a position to take full advantage of these opportunities.

**Conflicting Requirements**
The ideal basic spatial unit for both dwellings and business enterprises would be an address point representing actual physical location. The dilemma facing SNZ is that the spatial information needs of various organisations relating to household and business units are not consistent. While SNZ's interest is in the physical location of dwellings and enterprises, telecommunications and utility companies are more interested in the location of connections, while NZ Post is concerned with the physical location of the letterbox. It has been suggested that the Population Census might provide the opportunity for a complete GPS survey of all household locations in New Zealand. Physical location could even be recorded on the census enumeration maps. One approach might be to use the DCDB (Digital Cadastral Database) address points and land parcel centroids initially with a view to adjusting at a later date as the actual locations become known.

**Comparability**
While coding at meshblock level may still be the best option for population censuses, the question is often asked: should we be moving in a different direction with regard to meeting our business frame information needs? However, a major constraint binding any decision to move away from coding businesses at meshblock level is that this decision would reduce SNZ's ability to match business information with census information and thus to maximise the value of our information sources.

**Quality Issues**
Data is rarely provided at meshblock level in our major outputs. One reason for this has been a lack of assurance concerning the quality of coding at this level. However, the potential for improving meshblocking quality in rural areas is slowly being realised and implemented through a number of different initiatives. SNZ is now able to address the rural coding issues, via a combination of improved RAPID numbering as provided by territorial authorities and the impending development of an agricultural statistics frame which will link BF information to agricultural information on external databases[6]. The results of this investigation should also clarify a number of issues and alleviate many existing concerns.

**Confidentiality**
As the existing basic spatial unit, the meshblock is a central component of the BF and maintenance of the meshblock pattern is a core function within SNZ. There is a significant cost attached to this maintenance, including the work required annually to update the frame with new and changed meshblocks. If a new approach to the maintenance of spatial information were developed then the maintenance of the meshblock pattern would become less important, even unnecessary. If, for example, all statistics were geocoded to a point, then the only function of the meshblock would be to act as a standard unit for outputs to protect confidentiality. However confidentiality is a vital consideration and it is unlikely that SNZ would ever publish data at a lower level than meshblock. The use of the meshblock as both a storage and output unit offers built-in protection against loss of privacy. Since moving to a smaller unit increases the risk that individual details could be exposed, appropriate strategies would be needed. SNZ not only has a statutory requirement to protect privacy, it relies on the goodwill of the community for high response rates. Before other options can be considered, the issue of confidentiality of the data must be recognised and resolved.

Until then, the use of the meshblock coding pattern will continue to form the basis for representing spatial demographic characteristics of the NZ people and their business.

[6] Refer the progress report